DOCUMENT RESUME

ED 290 551 PS 017 096

AUTHOR Graue, M. Elizabeth; Shepard, Lorrie A.
TITLE Predictive Validity of the Gesell School Readiness Tests.
PUB DATE 88
NOTE 17p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Rating; *Elementary School Students; Grade 1; Kindergarten Children; *Predictive Validity; Primary Education; Referral; School Readiness; School Readiness Tests; *Screening Tests; *Test Validity
IDENTIFIERS *Developmental Testing; *Gesell School Readiness Test

ABSTRACT

In response to the fact that technical standards for screening and placement tests must be more rigorous than those for readiness tests, the predictive validity of the Gesell School Readiness Tests (GSRT) was examined. The purpose of the GSRT, a commonly used screening instrument, is the assessment of children's developmental behaviors to aid in placement decisions for young children. However, typical use of the Gesell test differs from most screening procedures in that it is not followed by a more comprehensive assessment. A sample of 45 first graders referred by their teachers for developmental testing and a random sample of 106 students were tested with the GSRT. Whether the test was administered as part of a normal referral process or as part of the special administration to a representative sample, each child's results were summarized as both a developmental age and a placement recommendation. Correlations were run on measured developmental age and student performance. A small positive relationship was found between Gesell developmental age and first grade report card grades. Additional outcome measures for a subgroup of the total sample indicated that the GSRT has modest predictive validity for standardized tests and low validity for teacher judgment of performance in first grade. Issues concerning misidentification of ready children and treatment efficacy are covered. It is concluded that the low predictive validity of the GSRT does not support its use for school readiness assessments leading to placement decisions. (RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED290551

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PREDICTIVE VALIDITY OF THE GESELL SCHOOL READINESS TESTS

M. ELIZABETH GRAUE
University of Colorado - Boulder

LORRIE A. SHEPARD
University of Colorado - Boulder

Predictive Validity of Gesell School Readiness Tests

Correspondence should be addressed to M. Elizabeth Graue, Campus Box 249, University of Colorado, Boulder CO 80309-0249. Phone (303) 492-8742.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
M. Elizabeth Graue
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PS 017096

BEST COPY AVAILABLE

ABSTRACT

The predictive validity of the Gesell School Readiness Tests is examined by correlating measured developmental age and performance in first grade. A sample of 45 students referred by their teachers for developmental testing and a random sample of 106 students chosen expressly for this study were tested with the GSRT. A small positive relationship was found between Gesell developmental age and first grade report card grades ($r = .23$). Additional outcome measures were available for a subgroup of the total sample and indicated that the GSRT has modest predictive validity for standardized tests and low validity for teacher judgement of performance in first grade. Issues concerning misidentification of ready children and treatment efficacy are also discussed. The low predictive validity of the GSRT does not support its use for school readiness assessments leading to placement decisions.

PREDICTIVE VALIDITY OF THE GESELL SCHOOL READINESS TESTS

Determining readiness for school experiences is a prevalent concern in early childhood education. Screening to identify children at risk has become common practice at both the preschool and kindergarten level. Provision of appropriate educational experiences and prevention of failure are often cited as the rationale for these screening programs, with screening instruments ranging from locally developed skills checklists to standardized batteries.

Meisels (1986) defines two types of tests. Developmental screening instruments "provide a brief assessment of the developmental abilities highly associated with children's future school success." Screening is intended as an initial step, possibly leading to more thorough assessment, for the purpose of identifying abnormal development and making special placements. Criteria for the selection of such tests include predictive validity, developmental content, and normative standardization. In contrast, school readiness tests "are concerned with which curriculum-related skills a child has already acquired." These tests should be criterion-referenced and the content should be consistent with the values and curricular approach embraced by the program the child is entering. Developmental screening instruments are useful for referral, leading to more thorough diagnostic assessment and special education placement decisions; while school readiness tests inform classroom instructional decisions. Meisels emphasizes that one type of test cannot be substituted for the other. The inappropriate use of screening instruments is compounded by lack of precise language to define the two types of instruments. Terms such as screening, readiness, and development are used in descriptions of both developmental screening and school readiness tests, making their purpose difficult to ascertain.

The focus of this study is to examine the predictive validity of a commonly used screening instrument, the Gesell School Readiness Tests. The expressed purpose of this test is the assessment of developmental behaviors to aid in placement decisions for young children. This purpose parallels the functions outlined by Meisels for developmental screening instruments. Typical use of the Gesell test differs from most screening procedures, however, because it is not followed by a more comprehensive assessment. Nonetheless, it is intended to measure developmental constructs rather than readiness skills and is used to make special placements such as developmental kindergarten and transition room. Predictive accuracy is of prime concern when a test is used for individual placement decisions because of the danger of misidentification and subsequent inappropriate special placement. More rigorous technical standards are held for screening or placement tests than for readiness tests because of the seriousness of the decisions made as the result of the test (APA, 1985).

The Gesell School Readiness Tests

Hundreds of school districts in the United States are currently using the Gesell School Readiness Tests to determine student placement. Testing may occur for all students during "Kindergarten Round-up" or on a referral basis during the kindergarten year. The tests reflect the philosophy of the Gesell Institute, which is based on Gesell's theory of maturational readiness. This theory states that behavior develops in predictable stages that are determined by a child's internal maturational clock.(Gesell Institute, 1982) The fact that progress through developmental stages is seen as immutable and internally controlled has two basic implications. The first is that environmental factors have relatively little impact on the rate of development. In fact, the main cause for failure among young children is purported to be inappropriate demands made on developmentally immature children. As we shall see, theoretical assumptions have implications for educational treatment.

The second implication of the maturational readiness theory is that developmental level can be measured through relative progress in the prescribed behavioral stages. This measure of development can then be used as an indication of school readiness. This is the purpose of the Gesell School Readiness Tests.

An individually administered test, the Gesell School Readiness Tests (GSRT) include the following tasks:

1) Initial Interview: The child is asked to give her name, birthday, the names and ages of her siblings, and father's occupation. Examiners are free to develop their own bank of questions, but are encouraged to use the same questions regularly so that they can make their own comparative decisions.

2) Paper and Pencil Tests: The child is asked to write her name, address, and numerals 1-20.

3) Copy Forms: The child is asked to copy a circle, cross, square, equilateral triangle, divided rectangle (a rectangle with lines that connect the corners and midpoints of the sides), and diamond. If successful in copying the 6 two-dimensional figures, the child older than 5 may attempt to draw a cylinder and a cube.

4) Incomplete Man: Presented with a partially drawn person, the child is asked to complete the missing facial features and body parts. In addition, the child is questioned about how the man feels.

5) Right and Left: The child is asked to name selected body parts, to identify her left and right hand, and to follow single (Touch your eye.) and double task commands (Touch your right thumb with your right little finger.)

6) Monroe Visual Tests: The child is asked to match pairs of designs or to reproduce complex designs from memory.

7) Naming Animals: The child is asked to name as many animals as she can within one minute.

8) Home and School Preferences: The child is asked to talk about what she likes to do best; and more specifically what she likes to do indoors and outdoors, both at home and at school.

In all cases, the examiner is to take into account both the content of response to a task and the manner of the response as well. Facial expression, pencil grip, and direction of drawing stroke are all included in scoring responses.

The results of the GSRT are used to make individual placement decisions based on an assessment of readiness for school experiences. The problem of lack of readiness is addressed by providing the child with time to develop outside the traditional school progress track. According to the Gesell Institute (1982), "the gift of time" can be provided through an extra year at home before kindergarten, an additional year in kindergarten or first grade, or in a transitional program between kindergarten and grade one.

Research on the GSRT

Very little technical information exists concerning the psychometric properties of the GSRT. Ames and Ilg (1978) report a correlation of .74 between the GSRT prediction of readiness and grade placement six years later. A correlation of .64 was found by Kaufman and Kaufman (1972) between the GSRT and the Stanford Achievement Test, administered in first grade. Wood, Powell and Knight (1984) obtained a 78% agreement rate between Gesell developmental age and teacher assessment of failure in kindergarten.

The Ames and Ilg and the Wood et al. studies suffer from limitations that erode the meaningfulness of their results. In both cases, the correlations reported are suspect as validity evidence to the extent that there was criterion contamination. In the Ames and Ilg study, the results of the GSRT were used to make placement decisions, then grade placement was used as the validity criterion. In the Wood study, the test followed the criterion (students were tested after their teachers determined that they had failed).

A second concern can be raised in both the Wood and the Kaufman findings. While seemingly high correlations are reported in each case, they do not signify accurate placement decisions. Shepard and Smith (1986) note that in the Kaufman study, a correlation of .64 translates into a standard error of measurement of six months, thus, a developmental age of 4.5 could not be distinguished from a score of 5.0. This breakpoint is often used in making a recommendation for kindergarten entrance. In the case of the Wood study, a seemingly high agreement rate is produced mostly by the successful children, correctly identified. When one looks at the children labeled at risk, however, for every potential kindergarten failure correctly identified, a successful child was incorrectly identified (Shepard & Smith, 1986).

Although predictive validity is the preeminent validity concern for tests used in selection decisions, predictive validity is but part of construct validation. An additional source of evidence to support the construct validity of an instrument is treatment efficacy related to the use of the instrument (Cronbach & Meehl, 1955). In the case of the GSRT, one would expect that students who received an alternative placement as a result of their performance on the GSRT would derive a benefit from that treatment compared to other children who did not receive a treatment matched to developmental level. According to the model of scientific theory building, evidence of treatment efficacy supports the validity of the theory, instrumentation, and treatment. While not directly related to the predictive validity of the test, the issue of treatment validity is extremely important, especially because the use of the GSRT leads to treatment in its recommendations.

In a study that addressed the concern about treatment validity, May and Welch (1984) compared students identified as developmentally mature (Traditional), students who were determined to be unready and who spent an additional year in school before placement in second grade (Buy-a-Year) and students who, though identified as unready, were promoted with their age group as a result of parent request (Over-placed). Despite one group being one year older and having had an extra year of school, no difference was found between the Buy-a-Year and the Over-placed students at the end of third grade on either a state administered achievement test or on the Stanford Achievement Test. In addition, when comparing the Over-placed students with the Traditional students, May and Welch again found no difference. The authors concluded that the Over-placed students, who would have appeared to be at risk according to their GSRT scores, had not experienced the predicted difficulties and that the Buy-a-Year group had not benefitted from the additional year spent in school. When a treatment makes no impact, it could either be because 1) the placement decisions are unreliable or 2) the treatment is ineffective. Studies of this type cast doubt on the test-treatment package but do not address directly the question of predictive validity.

Method

Subjects

Two samples were identified from a moderate size, middle class school district (20,000 students). During the 83-84 school year, a sample of 59 students had been referred by their teachers for developmental testing. Of this group, 34 were kindergarteners and 25 were first graders. In addition, a random sample of 125 kindergarten students was selected expressly for the predictive validity study. The referred sample was given the Gesell School Readiness Tests between October and April; the representative sample was tested by Gesell trained administrators

during May of the 1983-84 kindergarten year. The mean Developmental Ages obtained on the Gesell for the subgroups of the sample were:

Random Sample kindergarteners (May)	5.62
Referred Sample kindergarteners (October-April)	5.44
Referred Sample first graders (October-April)	5.89

Of the original 184, the 151 students included in this analysis continued to attend the school district at the end of their first grade year. Attrition reduced the size of the referred sample to 45 and the random sample to 106. Of these students, 123 were promoted to the next year with their age mates, 22 were retained in kindergarten (15 from the referred group, 7 from the random), and 6 were retained in grade 1 (all from the referred group of first graders).

Gesell Variables

Whether the test was administered as part of a normal referral process or as part of the special administration to a representative sample, each child's results were summarized as both a developmental age and as a placement recommendation. The developmental ages were indicated in the following ranges:

4.5	4.5 - 5	5	5 - 5.5	5.5 - 6
6	6 - 6.5	6.5	6.5 - 7	

Decision rules based on GSRT Developmental Age were applied according to school level criteria. In general, the cutoff points at the end of kindergarten were:

Below 5 to 5.5 years	Hold or Pass & Watch
5.5 years	Pass & Watch or Pass
Above 5.5 to 6 years	Pass

The recommendation of Pass & Special Education Referral was made for students with a D.A. of 4.5 to 5-5.5 years. The following recommendations were made, based on the students' GSRT performance:

Recommendation	Referred Sample K & 1		Random Sample K		Total Sample	
Retain in K	17	34%	17	14%	34	20%
Retain in 1st	16	31%			16	10%
K-1 Placement	2	4%	0	0%	2	1%
Pass to 1st & Special Education Referral	3	6%	0	0%	3	2%
Pass to 1st & Watch	8	15%	21	18%	29	17%
Pass to 2nd & Watch	3	6%			3	2%
Pass to 1st	2	4%	80	68%	82	48%
Pass to 2nd	0	0%			0	0%

Outcome Variables

Grade one report cards were used as the source of the dependent variables for the analysis. Grades in the following subject areas were coded on a scale from 1 to 5 (1=low and 5=high): Reading, Language, Math, Science, Social Studies, Work Habits, and Social Growth. An additional variable, Overall Grade, was created, to represent a global measure of student progress. For students who had spent two years in first grade, both years' data were collected.

Subgroups of the sample provided additional outcome measures. A different random sample of kindergarten classes in the district had been given the Metropolitan Readiness Tests in April, 1984. Forty one of the students in the Gesell sample participated in this program and therefore had pre-reading scores. Kindergarten students normally promoted to grade one had Comprehensive Test of Basic Skills (CTBS) scores at the end of first grade, given as part of the district testing program. Due to a change in the district-wide standardized testing program, first grade CTBS scores were not available the following year for students who had been retained in kindergarten. In Spring 1985, first grade teachers provided rankings of their students according to both grade level standards and relative standing within class in the following areas: Reading, Math, Social Maturity, Learner Self-Concept, and Appropriate Attention. Grade level ratings were on a four point scale: Above Grade Level, Grade Level, Below Grade Level, and Recommended to Repeat . Relative rankings were coded on a scale from one to five, with one being 'In the lowest 20%' of the class and 5 signifying 'In the highest 20%.' As in the case of CTBS scores, this information was not available for retained kindergarten students because it was collected only for the 84-85 school year.

Results

The means and standard deviations for first grade report card grades are presented in Table 1 for both the Random and Referred groups. For the six referred children who repeated first grade, data from the first year of first grade were used to avoid the confounding of the effect of retention with criterion performance. In each subject area but Language,the Random group had higher mean grades, with less variability than the Referred group.

Table 1

Means and Standard Deviations of First Grade Report Card Grades

	Referred Group			Random Group		
Subject Area	Mean	SD	n	Mean	SD	n
Reading	2.35	1.37	45	3.48	1.11	106
Language	2.76	1.21	45	2.43	.91	104
Math	2.96	1.37	45	3.29	.91	106
Science	3.07	.96	45	3.18	.53	105
Social Studies	3.02	.77	45	3.19	.54	105
Work Habits	3.17	.76	29	3.36	1.04	75
Social Growth	3.21	.99	28	3.29	1.07	80
Overall Grade	2.76	1.10	45	3.32	.76	106

Correlations of the GSRT with first grade report card grades are shown in Table 2. In all cases there is only slight evidence of a relationship between the GSRT and first grade performance. Weak, but significantly non-zero correlations are found for both the total group and the random sample in most subject areas. The GSRT and first grade outcome variables correlated poorly for the referred group, with none reaching statistical significance.

Because the Total sample includes a disproportionately large sample of the at risk (i.e. referred) population, its variance is unduly exaggerated. By definition, the random sample is representative of the total population and consequently should be the primary basis of interpretations. Note that the random sample is unrestricted on the report card grade criterion since later grade one grades were obtained even for those who were retained.

Table 2
Correlations of the Gesell School Readiness Test Developmental Age with First Grade Report Card Grades

	Gesell School Readiness Test Developmental Age		
	Total Sample n=151	Referred Sample n=45	Random Sample n=106
	r	r	r
Subject Grades:			
Reading	.16*	-.02	.21*
Language	.24*	.10	.29*
Math	.13	-.07	.25*
Science	.14	.08	.19*
Social Studies	.17*	.08	.22*
Work Habits	.23*	.09	.27*
Social Growth	.13	.26	.09
Overall Grade	.23*	.06	.31*

*$p < .05$

The correlations of the GSRT Developmental Age and additional outcome variables for subgroups of the total sample are presented in Table 3. With the exception of the Metropolitan Readiness Test scores, the students included in the analysis of this data were those promoted with their age mates and could therefore be considered a slightly restricted sample to the extent that poor performance on the GSRT influenced the decision to retain.

An additional sample subgroup is presented in Table 3 to address the issue of range restriction. The concern is that the exclusion of retained children reduced the variability of the sample and hence, unfairly weakened the validity of the correlations in Table 3. In an attempt to circumvent this dilemma, students in the random sample from schools with low kindergarten retention rates (0-4%) were analyzed separately. This sample can be seen as unrestricted because developmentally young students were not excluded, therefore the correlations have the benefit of the full range of both developmental age and first grade outcomes.

Table 3
Correlations of the GSRT Developmental Age
with Standardized Tests or Teacher Ratings

MEASURE:	TOTAL SAMPLE r	TOTAL SAMPLE n	REFERRED SAMPLE r	REFERRED SAMPLE n	RANDOM SAMPLE r	RANDOM SAMPLE n	NONRETAINING SCHOOLS RANDOM SAMPLE r	NONRETAINING SCHOOLS RANDOM SAMPLE n
Metropolitan Readiness Test Pre-reading Percentile	.40*	41	.36	7	.40*	34	.87*	5
CTBS Reading National Percentile	.40*	126	.59*	27	.34*	99	.44*	35
CTBS Math National Percentile	.40*	127	.31	28	.36*	99	.36*	35
Teacher Grade Level Ratings								
Reading Achievement	.19*	125	.25	23	.18	102	.14	32
Math Achievement	.28*	125	.45*	23	.24*	102	.12	31
Social Maturity	.19*	126	.37	23	.15	103	.12	32
Self Concept	.20*	126	.21	23	.20*	103	.21	32
Appropriate Attention	.20*	126	.20	23	.19*	103	.23	32
Teacher Relative Ratings								
Reading	.21*	126	.22	23	.21*	103	.21	32
Math	.28*	126	.42*	23	.25*	103	.16	32
Social Maturity	.23*	126	.52*	23	.17	103	.16	32
Self Concept	.16	126	.20	23	.16	103	.03	32
Appropriate Attention	.24*	126	.31	23	.22*	103	.18	32

*p < .05

The Metropolitan Readiness Test (MRT) could be seen as a type of concurrent validity measure, as it was also administered in the spring of the kindergarten year. The GSRT correlates moderately (r = .40) with this measure of first grade readiness. A much higher correlation (r = .87) is found for the unrestricted group but is based on only five students from a single school, one of whom had extremely lower scores on both measures. If that student's scores are removed, the correlation drops to .34. The MRT and the GSRT were developed for different purposes; the MRT for instructional planning based on academic readiness and the GSRT for placement decisions based on developmental age. Their underlying conceptions of readiness are quite different. Therefore it is not surprising that they are only modestly related.

The GSRT' s correlation with a standardized measure of first grade performance, the CTBS, is consistently modest but significantly non-zero ($r = .40$, $p < .05$) for the total sample of promoted students, with a larger correlation ($r = .59$, $p < .05$) between the GSRT and the CTBS Reading score for the Referred Sample. It is important to note that on most measures the Referred group is more varied, including possible Special Education referral, as well as some children considered by their teachers to be "bright-immature." This variability acts to inflate the correlation of Developmental Age with the other measures. There is little difference among the correlations for the unrestricted-random group and the full random sample, thus the most accurate coefficients to attend to are those from the random sample.

Low correlations ($r = .16$ to $.25$ in the random sample) were obtained between the GSRT score from kindergarten and Teacher ratings at the end of first grade.

The previous analyses do not take into account directly the fact that some students received an intervention by being retained in kindergarten. Theoretically, the success of retention prior to first grade would weaken the correlation by disrupting the original prediction. Whether one is concerned about range restriction or the confounding of treatment and prediction, the random sample, where only six students are missing and the low retaining group where no students are missing (i.e. none were retained) provide the most accurate picture of predictive validity.

A final set of data are presented in Table 4. These are within group correlations between Gesell Developmental Age and first grade grades. They are the same data reported in Table 2 but have been recomputed within-group depending upon whether children were retained or normally promoted. Because predictive validity refers to the accuracy of a test in distinguishing at-risk from normal children, the within group correlations do not reflect predictive validity, but could been seen as exploring the relationship between the test result and grades in subgroups that differ both by initial risk and treatment.

Table 4
Correlation of GSRT Developmental Age with First Grade Report Card Grades by Promotion Group

	REFERRED SAMPLE				RANDOM SAMPLE	
	Promoted	Retain K	Retain Grade 1	Retain Gr 1 2nd Year	Promoted	Retained K
	n=24	n=15	n=6	n=6	n=99	n=7
Subject:	r	r	r	r	r	r
Reading	.08	.20	.53	.14	.29*	-.49
Math	-.08	.57*	.41	-.06	.26*	-.22
Science	.01	.62*	.55	-.11	.19*	-.71
Social Studies	-.04	.62*	.55	-.11	.21*	.00
Work Habits	.13	-.19	.34	.55	.22	.86*
Social Growth	.42	-.04	.33	.33	.05	.76*
Overall	.10	.52*	.59	.00	.31*	-.51

*p < .05

Examining the correlations for the Referred Sample, there again appears to be very little relationship between the Developmental Age found by the GSRT in kindergarten and later first grade performance for all groups but those who were retained in kindergarten or first grade. Relatively strong and significant correlations were found in the areas of math, science, social studies, and overall grade performance for the kindergarten retainees. Ironically, high correlations for the retained kindergarten group after they received the treatment of an extra year is exactly where high correlations are least desirable, indicating that the extra year in kindergarten had done little to change the relative performance of these students. Those with a higher Developmental Age tended to obtain better grades, while those with lower Developmental Ages had lower grades even though they were all a year older than their first grade classmates and two years had passed since the assessment of Developmental Age. The first grade retained group is too small to interpret confidently, but the correlations are in the expected direction. Higher correlations are evidenced before the treatment and near zero correlations predominate post-treatment.

Weak correlations were found for the Random Sample Promoted group (n=99) in most subject areas. For Random Sample kindergarten retainees (n=7), strong and significant correlations were found for Work Habits and Social Growth, again indicating a lack of change in these variables even after an additional year in kindergarten. Although the size of this sample is very small, making inferences difficult, the negative correlations for this group are worrisome. They indicate that those with the highest Developmental Age in kindergarten received the worst

grades in first grade, after being retained. Combined with the data for the Referred retainees, the effectiveness of the retention treatment is questionable.

Strong within group correlations are exactly what one does not want if the test is to be used to distinguish groups. Only the drop off in correlations from year one to year two for the first grade sample follows the pattern that one would expect if the test were predicting accurately and the retention treatment were beneficial. Even here, the substantial correlations for work habits and social growth after an extra year are troublesome, suggesting that both the Gesell examiners and first grade teachers two years later might be attending to relatively enduring behavior patterns rather than immaturity intended to be measured by the Gesell. It should be noted that this same anomolous pattern occurred in the random sample kindergarten retainees but not in the referred kindergarten retainees.

Discussion

In general, it appears from our analysis that the Gesell School Readiness Tests are not potent predictors of first grade achievement. When related to a measure of first grade performance in the form of first grade report card grades, only a small positive relationship can be discerned. Data for students for whom no treatment was suggested (i.e. were judged to minimally at risk by the GSRT) indicate that the GSRT has modest predictive validity for standardized tests and low validity for teacher judgements of performance

In examining the correlations between the predictors and criteria, it is important to consider what a correlation of .20-.40 (the most prevalent in this study) signifies. Correlations in this range indicate very little shared variance between the predictor and the criterion and are therefore suspect for use in placement decisions. For example, Karl White (1976) found that across many studies, the typical correlation between socio-economic status (SES) and achievement is .25, but we rightly make no decisions using SES as a predictor.

In addition, classification error is a major concern when predictive validity coefficients are so small. For example, using a correlation of .23, as was obtained for the GSRT Developmental Age and overall grade for the total sample, and selecting the one-third who are least ready, only 41%* of those predicted to be at risk would in fact have problems later. As a result, 3 of 5 children identified as unready would actually be successful. In the case of a relatively high correlation, such as .59, 60% or roughly 3 of 5 would be correctly identified. It is a well known statistical phenomenon that even seemingly moderate predictive validities result in substantial misidentification, giving rise to great concern about their use for individual placement decisions.

---

*Calculation based on Taylor-Russell Tables of the Proportion Who Will be Satisfactory Among Those Selected

The value of the predictive contribution of an instrument is related to the benefit of the intervention that results from its interpretation. For example, when considering the validity of the identification procedures for the mildly retarded, a National Academy of Sciences Panel (Heller, Holtzman, & Messick, 1982) noted that if placement in special education were unambiguously a benefit, there would be less concern about misidentification. The treatment prescribed by the GSRT is more time to develop, in the form of retention in kindergarten or a transitional placement before grade 1. Neither intervention has been borne out as advantageous in studies on additional year programs (May & Welch 1984, Gredler 1984, Shepard & Smith 1985).

The low predictive validity of the GSRT, combined with questionable treatments related to the test's results, provide little evidence to advocate its use for placement decisions. Using Meisels' criteria for developmental screening, the GSRT lacks a primary component, predictive validity, and therefore does not meet the goals of the screening process. Its use could result in misidentification of a large number of students as unready. In typical samples of kindergarteners, more than half of the children predicted by the Gesell test to be unsuccessful in first grade would in fact be successful if they were allowed to be promoted normally.

References

American Psychological Association (1985). *Standards for educational and psychological tests and manuals.* Washington, D.C.

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302..

Gesell Institute of Human Development. (1982). *A gift of time.* New Haven, CT: Author.

Gredler, G.R. (1984). Transition classes: a viable alternative for the at-risk child? *Psychology in the Schools, 21,* 463-470.

Heller, K.A., W.H. Holtzman, & S. Messick, eds. (1982). *Placing children in special education: a strategy for equity.* Washington, D.C.: National Academy Press.

Ilg, F.L., L.B. Ames, J. Haines & C. Gillespie (1978). *School readiness: behavior tests used at the Gesell Institute.* New York: Harper & Row.

Kaufman, A.S. & N.L. Kaufman (1972). Tests built from Piaget's and Gesell's tasks as predictors of first-grade achievement. *Child Development, 43,* 521-535.

May, D.C. & E.L. Welch (1984). The effects of developmental placement and early retention on children's later scores on standardized tests. *Psychology in the Schools, 21,* 381-385.

Meisels, S.J. (1986). Testing four- and five-year olds: response to Salzer and to Shepard and Smith. *Educational Leadership, 44,* 90-92.

Shepard, L. & M.L. Smith. (1985). *The Boulder Valley kindergarten study: retention practices and retention effects.* Boulder, CO: Boulder Valley Public Schools.

Shepard, L. & M.L. Smith (1986). Synthesis of research on school readiness and kindergarten retention. *Educational Leadership, 44,* 78-86.

Taylor, H.C. & J.T. Russell (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology, 23,* 565-578.

White, K.R. (1976). *The relationship between socio-economic status and academic achievement..* Doctoral dissertation, University of Colorado.

Wood, C., S. Powell, & R.C. Knight (1984). Predicting school readiness: the validity of developmental age. *Journal of Learning Disabilities, 17,* 8-11.